

Listening to the Universe with Gravitational-Wave Astronomy

Scott A. Hughes

*Kavli Institute for Theoretical Physics, University of California, Santa Barbara,
CA 93106*

*Department of Physics, Massachusetts Institute of Technology, Cambridge, MA
02139*

sahughes@mit.edu

The LIGO (Laser Interferometer Gravitational-Wave Observatory) detectors have just completed their first science run, following many years of planning, research, and development. LIGO is a member of what will be a worldwide network of gravitational-wave observatories, with other members in Europe, Japan, and — hopefully — Australia. Plans are rapidly maturing for a low frequency, space-based gravitational-wave observatory: LISA, the Laser Interferometer Space Antenna, to be launched around 2011. The goal of these instruments is to inaugurate the field of *gravitational-wave astronomy*: using gravitational-waves as a means of listening to highly relativistic dynamical processes in astrophysics. This review discusses the promise of this field, outlining why gravitational waves are worth pursuing, and what they are uniquely suited to teach us about astrophysical phenomena. We review the current state of the field, both theoretical and experimental, and then highlight some aspects of gravitational-wave science that are particularly exciting (at least to this author).

1 Motivation

The current state of gravitational-wave science is very similar to the state of neutrino science circa 1950 [1]: we have a mature theoretical framework describing this form of radiation; we have extremely compelling *indirect* evidence of the radiation's existence; but an unambiguous direct detection has not yet happened. Unlike the case of neutrinos, however, it is unlikely that a bright laboratory source of gravitational radiation (analogous to the Savannah River nuclear reactor) will be constructed (though see [2] for an alternative

view). The only guaranteed sources of gravitational waves bright enough to be measurable will arise from violent astrophysical events. Though perhaps somewhat frustrating on the one hand — we must remain patient while we wait for nature to supply us with a radiation source bright enough for our fledgling detectors — it offers a great opportunity on the other. Gravitational radiation promises to open a unique window onto astrophysical phenomena that may teach us much about “dark” processes in the universe. Once these detectors have met their “physics goal” of directly and unambiguously detecting gravitational waves, they will grow into observatories that — we hope! — will be rich sources of data on violent astrophysical events.

The properties of gravitational radiation and the processes that drive its emission are quite different from the properties and processes relevant to electromagnetic radiation. Consider the following differences:

- Electromagnetic waves are oscillations of electric and magnetic fields that propagate through spacetime. Gravitational waves are oscillations of spacetime itself. Formally, this is an extremely important difference, and historically has been a source of some controversy regarding the validity of certain computation schemes in gravitational-wave theory (with some members of the relativity community worrying that analogies to electromagnetic radiation were used without sufficient justification). This difference can make it difficult to define what exactly a gravitational wave *is*. One must identify an oscillating contribution to the curvature of spacetime that varies on a lengthscale $\lambda/2\pi$ much shorter than the lengthscales over which all other important curvatures vary. In this sense, gravitational waves are more similar to waves propagating over the ocean’s surface (varying on a lengthscale much smaller than the Earth’s radius of curvature) than they are to electromagnetic radiation.
- Astrophysical electromagnetic radiation typically arises from the incoherent superposition of waves produced by many emitters (e.g., electrons in the solar corona, hot plasma in the early universe). This radiation directly probes the thermodynamic state of a system or an environment. Gravitational waves are coherent superpositions arising from the bulk dynamics of a dense source of mass-energy. These waves directly probe the dynamical state of a system.
- Electromagnetic waves interact strongly with matter; gravitational waves do not. This follows directly from the relative strength of the electromagnetic and gravitational interactions. The weak interaction strength of gravitational waves is both blessing and curse: it means that gravitational waves propagate from emission to observers on the Earth with essentially zero absorption, making it possible to probe astrophysics that is hidden or dark — e.g., the coalescence and merger of black holes, the collapse of a stellar core, the dynamics of the early universe. This also means that the waves interact very weakly with detectors, necessitating a great deal of effort to

ensure their detection. Also, because many of the best sources are hidden or dark, they are very poorly understood today — we know very little about what are likely to be some of the most important sources of gravitational waves.

- The direct observable of gravitational radiation is the waveform h , a quantity that falls off with distance as $1/r$. Most electromagnetic observables [3] are some kind of energy flux, and so fall off with a $1/r^2$ law. This means that relatively small improvements in the sensitivity of gravitational-wave detectors can have a large impact on their science: doubling the sensitivity of a detector doubles the distance to which sources can be detected, increasing the volume of the universe to which sources are measurable by a factor of 8. Every factor of two improvement in the sensitivity of a gravitational-wave observatory should increase the number of observable sources by about an order of magnitude.
- Electromagnetic radiation typically has a wavelength smaller than the size of the emitting system, and so can be used to form an image of the source, exemplified by the many beautiful images observatories have provided over the years. By contrast, the wavelength of gravitational radiation is typically comparable to or larger than the size of the radiating source. Gravitational waves *cannot* be used to form an image. Instead, gravitational-waves are best thought of as analogous to sound: the two polarizations carry a stereophonic description of the source’s dynamics. Many researchers in gravitational-wave physics illustrate their work by playing audio encodings of expected gravitational-wave sources and of detector noise. Some source examples from this author’s research can be found at [4]; I leave it to the reader to judge whether they are beautiful or not.
- In most cases, electromagnetic astronomy is based on deep imaging of small fields of view: observers obtain a large amount of information about sources on a small piece of the sky. Gravitational-wave astronomy, by contrast, will be a nearly all-sky affair: gravitational-wave detectors have nearly 4π steradian sensitivity to events over the sky. A consequence of this is that their ability to localize a source on the sky is not good by usual astronomical standards; but, it means that any source on the sky will be detectable, not just sources towards which the detector is “pointed”. The contrast between the all-sky sensitivity but poor angular resolution of gravitational-wave observatories, and the pointed, high angular resolution of telescopes is very similar to the angular resolution contrast of hearing and sight, strengthening the useful analogy of gravitational waves with sound.

These differences show why we believe that gravitational-wave astronomy will open a radically new observational window for astrophysics, and motivate the efforts to construct sensitive gravitational-wave detectors. The last two points in particular explain why we have chosen to describe gravitational-wave astronomy as “listening to the universe”. (Marcia Bartusiak similarly expanded on this theme in her very engaging book “Einstein’s Unfinished

Symphony” [5].) Gravitational-wave astrophysics can be thought of as learning to speak the language of gravitational-wave sources so that we can understand and learn about the sources that the new detectors will measure.

This article surveys the current state of this field. Sections 2 and 3 are review material — Sec. 2 discusses the major background concepts associated with gravitational radiation and gravitational-wave detectors, and Sec. 3 surveys astrophysical sources and detection methods, categorizing them by the frequency band in which they primarily radiate. We then focus on several aspects of gravitational-wave astronomy involving black holes that are of particular interest to this author. Section 4 discusses the importance of binary black hole systems as sources of gravitational waves, and what can be learned from such observations from the standpoint of astrophysics and physics generally. Section 5 discusses in detail a special kind of binary black hole system — extreme mass ratio binaries, in which one black hole in the binary is far more massive than the other. We discuss the particularly powerful and interesting analyses that measurement of these waves can make possible, and then review the challenges that must be overcome to understand the language of these sources.

2 Major concepts of gravitational-wave physics

The idea that radiation of some sort might be associated with the gravitational interaction has a surprisingly long pedigree. As early as 1776, Laplace [6] suggested that an apparent secular acceleration in the Moon’s orbit (deduced by Edmund Halley from a study of medieval solar eclipses recorded by Al-Batanni and of still older eclipses recorded by Ptolemy [7]) could be explained by requiring that the gravitational interaction propagate at finite speed. (The correct explanation of this effect turned out to be tidal transfer of the Earth’s rotational angular momentum to the Moon’s orbit [7].) Poincaré somewhat tentatively resurrected this idea in 1908 in an attempt to explain the anomalous perihelion shift of Mercury [8]. (This effect was eventually explained by the nonlinear “post-Newtonian” effect of relativistic gravity [9].)

Gravitational waves finally and (almost) unambiguously entered the lexicon of physics as a natural consequence of general relativity. Soon after general relativity was introduced, Einstein predicted the existence of gravitational waves in a 1916 paper [10]. This analysis was flawed by a few important algebraic errors, which were corrected in a 1918 paper [11]. Einstein showed that gravitational radiation arises from variations in a source’s quadrupole moment, and derived (with a factor of 2 error) what has come to be called the “quadrupole formula” for the rate at which the radiation carries energy away from the source. This is what one expects intuitively — gravitational waves

arise from the acceleration of masses in a manner similar to the generation of electromagnetic radiation from the acceleration of charges. At lowest order, electromagnetic waves come from the time changing charge dipole moment, and are thus dipole waves; monopole EM radiation would violate charge conservation. We expect (at lowest order) gravitational waves to come from the time changing quadrupolar distribution of mass and energy, since monopole gravitational waves would violate mass-energy conservation, and dipole waves would violate momentum or angular momentum conservation.

The parenthetical “almost” at the beginning of the preceding paragraph refers to a rather lengthy controversy over the formal underpinnings of gravitational radiation calculations. These controversies mostly came to an end in the 1980s, thanks in large part to the careful, rigorous calculations of Thibault Damour and collaborators (cf. Ref. [12] and references therein) and the excellent correspondence to observations of the Hulse-Taylor binary pulsar [13,14]; see Ref. [7] for extended discussion. It is now generally accepted that Einstein’s original quadrupole formula (corrected for the factor of 2 error) properly describes at lowest order the energy flow from a radiating source (even if that source has strong self gravity, a major issue contributing to the aforementioned controversy), and we are likewise confident that theory can go well beyond this lowest order (see, e.g., the review by Blanchet [15] and references therein).

Gravitational waves act tidally, stretching and squeezing any object that they pass through. Their quadrupolar character means that they squeeze along one axis while stretching along the other. When the size of the object that the wave acts upon is small compared to the wavelength (as is the case for LIGO), forces that arise from the two GW polarizations act as in Fig. 1. The polarizations are named “+” (plus) and “ \times ” (cross) because of the orientation of the axes associated with their force lines.

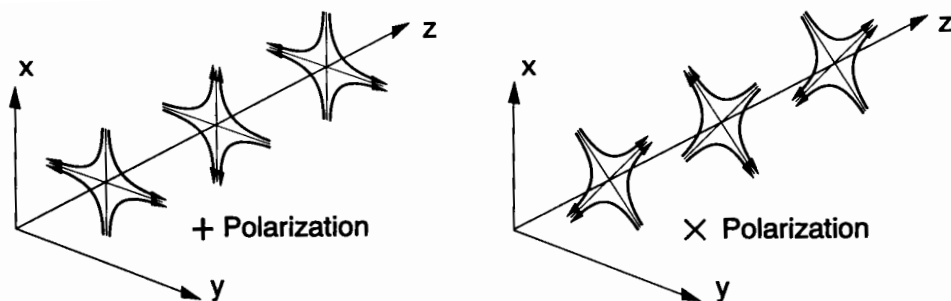


Fig. 1. The lines of force associated with the two polarizations of a gravitational wave (from Ref. [17]).

Interferometric gravitational-wave detectors measure this tidal field by observing their action upon a widely-separated set of test masses. In ground-based interferometers, these masses are arranged as in Fig. 2. The space-based de-

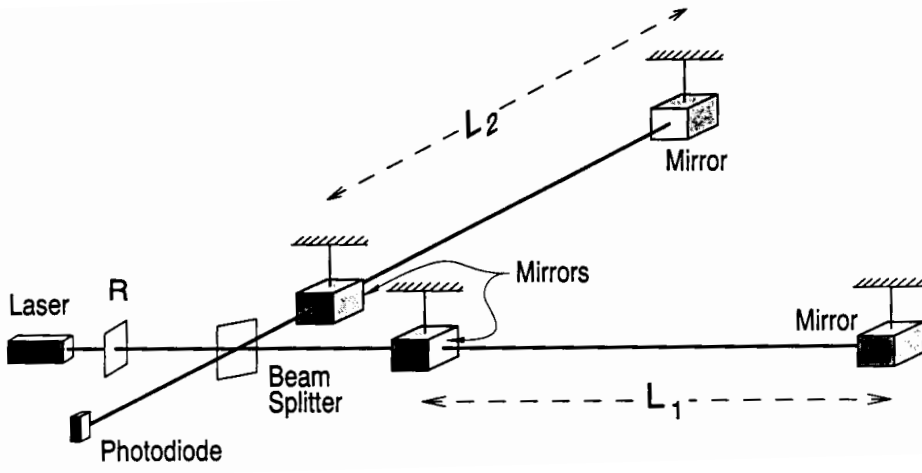


Fig. 2. Layout of an interferometer for detecting gravitational waves (from Ref. [17]).

tector LISA arranges its test masses in a large equilateral triangle that orbits the sun, illustrated in Fig. 3. On the ground, each mass is suspended with a sophisticated pendular isolation system to eliminate the effect of local ground noise. Above the resonant frequency of the pendulum (typically of order 1 Hz), the mass moves freely. (In space, the masses are actually free floating.) In the absence of a gravitational wave, the sides L_1 and L_2 shown in Fig. 2 are about the same length L .

Suppose the interferometer in Fig. 2 is arranged such that its arms lie along the x and y axes of Fig. 1. Suppose further that a wave impinges on the detector down the z axis, and the axes of the $+$ polarization are aligned with the detector. The tidal force of this wave will stretch one arm while squeezing the other; each arm oscillates between stretch and squeeze as the wave itself oscillates. The wave is thus detectable by measuring the separation between the test masses in each arm and watching for this oscillation. In particular, since one arm is always stretched while the other is squeezed, we can monitor the difference in length of the two arms:

$$\delta L(t) \equiv L_1(t) - L_2(t) . \quad (1)$$

For the case discussed above, this change in length turns out to be the length of the arm times the $+$ polarization amplitude:

$$\delta L(t) = h_+(t)L . \quad (2)$$

The gravitational wave acts as a strain in the detector; h is often referred to as the “wave strain”. Note that it is a dimensionless quantity. Equation (2) is easily derived by applying the equation of geodesic deviation to the separation of the test masses and using a gravitational-wave tensor on a flat background

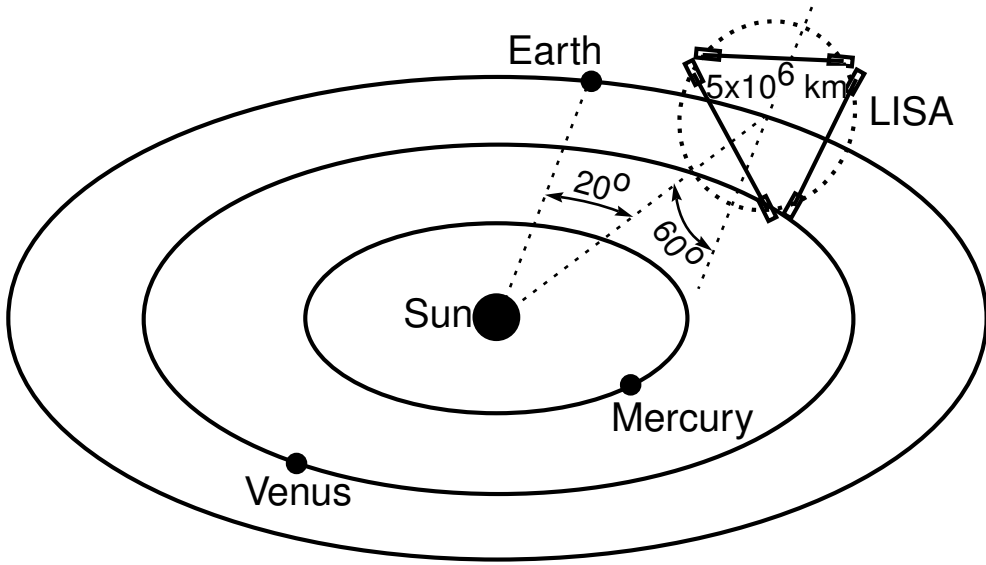


Fig. 3. Orbital configuration of the LISA antenna.

spacetime to develop the curvature tensor; see Ref. [18], Sec. 9.2.2 for details.

We obviously do not expect astrophysical gravitational-wave sources to align themselves in as convenient a manner as described above. Generally, both polarizations of the wave influence the test masses:

$$\frac{\delta L(t)}{L} = F^+ h_+(t) + F^\times h_\times(t) \equiv h(t) . \quad (3)$$

The antenna response functions F^+ and F^\times weight the two polarizations in a quadrupolar manner as a function of a source's position and orientation relative to the detector; see [18], Eqs. (104a,b) and associated text.

The energy flux carried by gravitational waves scales as \dot{h}^2 (where the over-dot denotes a time derivative). In order for the energy flowing through large spheres to be conserved, h must fall off with distance as $1/r$. As discussed above, the lowest order contribution to the waves arises from changes in a source's quadrupole moment. To order of magnitude, this moment is given by $Q \sim (\text{source mass})(\text{source size})^2$. By dimensional analysis, we then know that the wave strain must have the form

$$h \sim \frac{G}{c^4} \frac{\ddot{Q}}{r} . \quad (4)$$

The second time derivative of the quadrupole moment is given approximately by $\ddot{Q} \simeq 2Mv^2 \simeq 4E_{\text{kin}}^{\text{ns}}$; v is the source's internal velocity, and $E_{\text{kin}}^{\text{ns}}$ is the nonspherical part of its internal kinetic energy. Strong sources of gravitational radiation are sources that have strong non-spherical dynamics — for example,

compact binaries (containing white dwarfs, neutron stars, and black holes), mass motions in neutron stars and collapsing stellar cores, the dynamics of the early universe.

Violent events that are likely to be interesting gravitational-wave sources are very rare — for example, supernovae from the collapse of massive stellar cores appear to occur in our galaxy once every few centuries. For our detectors to have a realistic chance of measuring observable events, they must be sensitive to sources at rather large distances. For example, to have an interesting shot at measuring the coalescence of binary neutron star systems, we need to reach out to several hundred megaparsecs (*i.e.*, a substantial fraction of 10^9 light years) [19–21]. For such coalescences, $E_{\text{kin}}^{\text{ns}}/c^2 \sim 1$ solar mass ($\equiv 1 M_{\odot}$). Plugging into Eq. (4) gives the estimate

$$h \sim 10^{-21} - 10^{-22} . \quad (5)$$

This sets the sensitivity required to measure gravitational waves. Combining this scale with Eq. (3) tells us that for every kilometer of baseline L we need to be able to measure a distance shift δL of better than 10^{-16} centimeters.

This is usually the point at which people decide that gravitational-wave scientists aren’t playing with a full deck. How can we possibly hope to measure an effect that is $\sim 10^{12}$ times smaller than the wavelength of visible light? For that matter, how is it possible that thermal motions do not wash out such a tiny effect?

That such measurement is possible with laser interferometry was analyzed thoroughly and published by Rainer Weiss in 1972 [22]. (It should be noted that the possibility of detecting gravitational waves with laser interferometers has an even longer history, reaching back to Pirani in 1956 [23], and has been independently invented by Gertsenshtein and Pustovoit in 1962 [24] and Weber in the 1960s (unpublished), prior to Weiss’s detailed analysis. See Sec. 9.5.3 of Ref. [18] for further discussion.) Examine first how a laser with a wavelength of 1 micron can measure a 10^{-16} cm displacement. In a laser interferometer like LIGO, the basic optical layout is as sketched in Fig. 2. A carefully prepared laser state is split at the beamsplitter and sent into the Fabry-Perot arm cavities of the detector. The reflectivities of the mirrors in these cavities are chosen such that the light bounces roughly 100 times before exiting the arm cavity (that is, the finesse \mathcal{F} of the cavity is roughly 100). This corresponds to about half a cycle of a 100 Hz gravitational wave. The phase shift acquired by the light during those 100 round trips is

$$\Delta\Phi_{\text{GW}} \sim 100 \times 2 \times \Delta L \times 2\pi/\lambda \sim 10^{-9} . \quad (6)$$

This phase shift can be measured provided that the shot noise at the photo-

diode, $\Delta\Phi_{\text{shot}} \sim 1/\sqrt{N}$, is less than $\Delta\Phi_{\text{GW}}$. N is the number of photons accumulated over the measurement; $1/\sqrt{N}$ is the phase fluctuation in a quantum mechanical coherent state that describes a laser. We therefore must accumulate $\sim 10^{18}$ photons over the roughly 0.01 second measurement, translating to a laser power of about 100 watts. In fact, as was pointed out by Ronald Drever [25], one can use a much less powerful laser: even in the presence of a gravitational wave, only a tiny portion of the light that comes out of the interferometer's arms goes to the photodiode. The vast majority of the laser power is sent back to the laser. An appropriately placed mirror bounces this light back into the arms, *recycling* the light. The recycling mirror is shown in Fig. 2, labeled "R". With it, a laser of ~ 10 watts drives several hundred watts of input to the interferometer's arms.

Thermal excitations are overcome by averaging over many many vibrations. For example, the atoms on the surface of the interferometers' test mass mirrors oscillate with an amplitude

$$\delta l_{\text{atom}} = \sqrt{\frac{kT}{m\omega^2}} \sim 10^{-10} \text{ cm} \quad (7)$$

at room temperature T , with m the atomic mass, and with a vibrational frequency $\omega \sim 10^{14} \text{ s}^{-1}$. This amplitude is huge relative to the effect of gravitational radiation — how can we possibly hope to measure the wave? The answer is that atomic vibrations are random and incoherent. The ~ 7 cm wide laser beam averages over about 10^{17} atoms and at least 10^{11} vibrations per atom in a typical measurement. The effect is thus suppressed by a factor $\sim \sqrt{10^{28}}$ — atomic vibrations are *completely* irrelevant compared to the coherent effect of a gravitational wave. Other thermal vibrations, however, are not irrelevant and in fact dominate LIGO's noise in certain frequency bands. For example, the test masses' normal modes are thermally excited. The typical frequency of these modes is $\omega \sim 10^5 \text{ s}^{-1}$ and they have mass $m \sim 10 \text{ kg}$, so $\delta l_{\text{mass}} \sim 10^{-14} \text{ cm}$. This, again, is much larger than the effect we wish to observe. However, the modes are very high frequency, and so can be averaged away provided the test mass is made from material with a very high quality factor Q — the mode's energy is confined to frequencies near ω and doesn't leak into the band we want to use for measurements. Understanding the physical nature of noise in gravitational-wave detectors is an active field of current research; see Refs. [26–33] and references therein for a glimpse of recent work. In all cases, the fundamental fact to keep in mind is that a gravitational wave acts *coherently*, whereas noise acts *incoherently*, and thus can be beaten provided one is able to average away the incoherent noise sources.

3 Gravitational-wave frequency bands and measurement

It is useful to categorize gravitational-wave sources (and the methods for detecting their waves) by the frequency band in which they radiate. Broadly speaking, we may break the gravitational-wave spectrum into four rather different bands: the *ultra low frequency* band, $10^{-18} \text{ Hz} \lesssim f \lesssim 10^{-13} \text{ Hz}$; the *very low frequency* band, $10^{-9} \text{ Hz} \lesssim f \lesssim 10^{-7} \text{ Hz}$; the *low frequency* band, $10^{-5} \text{ Hz} \lesssim f \lesssim 1 \text{ Hz}$; and the *high frequency* band, $1 \text{ Hz} \lesssim f \lesssim 10^4 \text{ Hz}$.

For compact sources (mass/energy configurations that are of compact support), the band in which gravitational waves are generated is typically related to the source’s size R and mass M . R is meant to set the scale over which the source’s dynamics vary; for example, it could be the actual size of a particular body, or the separation of members of a binary. The “natural” gravitational-wave frequency of such a source is $f_{\text{GW}} \sim (1/2\pi)\sqrt{GM/R^3}$. Because $R \lesssim 2GM/c^2$ (the Schwarzschild radius of a mass M), we can estimate an upper bound for the frequency of a compact source:

$$f_{\text{GW}}(M) < \frac{1}{4\sqrt{2}\pi} \frac{c^3}{GM} \simeq 10^4 \text{ Hz} \left(\frac{M_{\odot}}{M} \right). \quad (8)$$

This is a rather hard upper limit, since many interesting sources are quite a bit larger than $2GM/c^2$, or else evolve through a range of sizes before terminating their emission at $R \sim 2GM/c^2$. Nonetheless, this frequency gives some sense of the types of compact sources that are likely to be important in each band — high frequency compact sources are of stellar mass (several solar masses); low frequency compact sources are of thousands to millions of solar masses, or else contain widely separated stellar mass bodies; etc. Other interesting sources of waves, particularly in the lower frequency bands, are not well-described by these compact body rules; we will discuss them separately in greater depth below.

3.1 High frequency

The high frequency band, $1 \text{ Hz} \lesssim f \lesssim 10^4 \text{ Hz}$, is the band targeted by the new generation of ground-based laser interferometric detectors, such as LIGO. (It also corresponds roughly to the audio band of the human ear: when converted to sound, LIGO sources are human audible without any frequency scaling.) The low frequency end of this band is set by the fact that it is extremely difficult to isolate against ground vibrations at low frequencies, and probably impossible to isolate against gravitational coupling to ground vibrations, human activity, and atmospheric motions [31–33]. The high end of the band